ELSEVIER

Discrete Applied Mathematics 65 (1996) 157-166

DISCRETE APPLIED MATHEMATICS

On preemptive scheduling: a general setting for the two-phase method

M. Cochand[a, c, *], A. Gaillard[b, c]

[a] Institute of Mathematics, University of Lausanne, Lausanne, Switzerland

[b] Institute of Informatics, University of Fribourg, Fribourg, Switzerland

[c] Institute of Operations Research, Swiss Federal Institute of Technology, CH-8092 Zürich, Switzerland

Received 11 January 1993; revised 24 August 1993

Abstract

The validity of the two-phase method for preemptive scheduling is established in a wide context by means of a classical result of polarity. A geometrical interpretation is given and relations to minimal length almost nonpreemptive scheduling are discussed in this general setting.

1. Introduction

The object under study in this paper is the following linear program:

$$(LP) \qquad \begin{cases} t_0 = \min t, \\ t \geqslant \lambda c, \\ \lambda A = x^{\mathrm{T}}, \\ Bx + td \leqslant b, \\ \lambda \geqslant \mathbf{0}, \end{cases}$$

with variables (x, t, λ), where $A \in \mathbb{R}^{s \times p}$ and $B \in \mathbb{R}^{n \times p}$ are matrices; $\mathbf{0} < c \in \mathbb{R}^s_+$, $d, b \in \mathbb{R}^n$, $x \in \mathbb{R}^p$ are column vectors; $\lambda \in \mathbb{R}^s$ is a row vector, and $t \in \mathbb{R}$.

The reason for studying such a linear program LP is to provide a common framework for several known results addressing preemptive scheduling problems which can naturally be formulated in the above form. To make this idea more precise,

* Corresponding author. Corresponding address: Institute of Operations Research, Swiss Federal Institute of Technology, CH-8092 Zürich, Switzerland.

SSDI 0166-218X(95)00033-X

let us first recall the main features of preemptive scheduling as well as the so-called two-phase method for solving special cases of it.

Preemptive scheduling deals with a set of jobs $\boldsymbol{J} = \{J_1, \dots, J_n\}$ which have to be processed by means of a set of processors $\boldsymbol{P} = \{P_1, \dots, P_m\}$. Preemption is allowed, that is, the execution of a job on a processor can be interrupted without penalty at any time and continued, later or immediately, on any processor able to process it (according nevertheless to certain rules specifying what can be processed simultaneously). An instance of this problem is defined by three types of data:

(a) Speed of execution: for each processor P_i and job J_j we are given the time p_{ij} which would be required to process completely J_j on P_i.

(b) Simultaneity of processing: call *action* the fact of executing part of a job J_j on a processor P_i, so that an action is determined by a pair (i, j). An instance of the preemptive scheduling problem is characterized by those sets of actions which can occur simultaneously. Such a set will be called an *operating mode* and will be represented by its $\{0, 1\}$-incidence vector in $\{0, 1\}^{P \times J} = \{0, 1\}^{mn}$. Any matrix $A \in \{0, 1\}^{s \times mn}$ whose rows A_i are the incidence vectors of the operating modes will be called the *matrix of operating modes* of the given instance.

(c) Cost of processing: for each operating mode A_i a cost rate $c_i > 0$ for applying A_i during one time unit is given.

The goal considered here is to determine a schedule of minimal cost for the processing of all jobs.

A schedule is completely determined by a sequence $(\lambda_k,\ A_{g(k)})_{k=1}^{s}$ which describes times λ_k during which operating modes $A_{g(k)}$ will be applied, in the order given by the sequence. The cost of the schedule is $\sum_{k=1}^{s} c_k \lambda_k$.

This preemptive scheduling problem can be expressed as a linear program LP' of the form LP, where A is the matrix of operating modes, the component x_{ij} of vector x is the time spent by job J_j on processor P_i in the schedule determined by the vector λ, and $Bx + td \leqslant b$ stands for the set of constraints $\sum_{i=1}^{m} x_{ij}/p_{ij} = 1$ $(j = 1, \dots, n)$ which guarantees the completion of job J_j ($d = \mathbf{0}$ in this case).

For the special case, where the objective is to find a schedule of minimal length (i.e. $c_i = 1$ for all A_i's) and where furthermore the family of operating modes is implicitly given by the rule:

> *A job cannot be on more than one processor at a time and similarly no processor can work on more than one job at a time,*

the problem has been solved by Slowinski and Weglarz [13] and independently by Lawler and Labetoulle [10]. (We shall refer to this situation as the *basic model* and denote by LP'' the corresponding linear program of the form LP'.) These authors showed that in such a situation, the problem can efficiently be solved by the so-called *two-phase method* which runs as follows:

Phase 1: Determine the duration t^* and the vector x^* of total times x_{ij} jobs J_j's are processed on the P_i's in an optimal solution of LP''. This pair (x^*, t^*) is obtained as an optimal solution of the following linear program $LP1''$:

$$(LP1'')\quad \begin{cases} \min t, & \\ \sum_{j=1}^{n} x_{ij} \leqslant t, \quad i = 1, \dots, m, & (1.1) \\ \sum_{i=1}^{m} x_{ij} \leqslant t, \quad j = 1, \dots, n, & (1.2) \\ Bx \leqslant b, & \\ x_{ij} \geqslant 0, \quad i = 1, \dots, m, & \\ \qquad\quad\ \ j = 1, \dots, n. & \end{cases}$$

Notice that $LP1''$ depends only on the variables x and t, and not on λ.

Phase 2: Solve the partitioning problem $\lambda A = x^{*\mathrm{T}}$, $\lambda \geqslant \mathbf{0}$, by a vector λ^* satisfying $\lambda^* \mathbf{1} = t^*$ (so that (x^*, t^*, λ^*) is an optimal solution of LP'').

The validity of a similar two-phase method for some families of operating modes extending the basic model of [10, 13] has been shown by Slowinski [12] and de Werra [5, 6]. These authors exhibited a linear program for Phase 1 and a partitioning algorithm for Phase 2, which is used "a posteriori" to prove the optimality of the result of Phase 1.

The question underlying this work is: Is it possible to solve the linear program LP (and therefore the general preemptive scheduling problem LP') by a generalized two-phase method? More precisely, does a linear program $LP1$ exist which depends only on the variables x and t, and not on λ, such that an optimal solution (x^*, t^*, λ^*) of LP can be found by the following two phases:

Phase 1: Determine (x^*, t^*) as an optimal solution of $LP1$.

Phase 2: Determine λ^* as a solution of $\min \lambda^* c$ s.t. $\lambda A = x^{*\mathrm{T}}$, $\lambda \geqslant \mathbf{0}$.

This question will be settled affirmatively.

The organization of the paper is as follows. In the next section we give our main result, that problem LP can always be solved in two phases. We then give a geometrical interpretation of the (generalized) two-phase method which makes clear how and why the determination of t^* (and x^*) can occur prior to the one of the actual schedule determined by λ^*. In Section 3 we show how previous preemptive scheduling results fit in our setting and discuss how resources can be taken into account. In the last section the applicability of the two-phase approach is considered for some cases of almost nonpreemptive scheduling.

Throughout the paper we will use the following notations. Unless otherwise specified, a vector $x \in \mathbb{R}^n$ is a column vector and x^{T} denotes its transpose. For $k \in \mathbb{R}$, $\boldsymbol{k}$ denotes the vector (row or column) of dimension appropriate to the context and having all its components equal to k. Finally, for a given matrix A, $Conv(A \cup \{\mathbf{0}\})$ (resp. $Cone(A)$) stands for the convex hull of the rows A_i of A and the vector $\mathbf{0}$ (resp. the cone of the rows A_i of A).

2. General form of the two-phase method

Let A be the matrix and c the vector of program LP and $\bar{A} \in \mathbb{R}^{u \times p}$, $\bar{R} \in \mathbb{R}^{v \times p}$ be matrices such that

$$P := Conv(\{1/c_i \cdot A_i \mid 1 \leqslant i \leqslant s\} \cup \{\mathbf{0}\}) = \{x \in \mathbb{R}^p \mid \bar{A}x \leqslant \mathbf{1},\ \bar{R}x \leqslant \mathbf{0}\}. \quad (2.1)$$

Consider for B, d and b given in LP the following linear program $LP1$ with variables $t \in \mathbb{R}$ and $x \in \mathbb{R}^p$:

$$(LP1) \qquad \begin{cases} t_1 = \min t, \\ \bar{A}x \leqslant t\mathbf{1}, \\ \bar{R}x \leqslant \mathbf{0}, \\ Bx + td \leqslant b, \\ t \geqslant \mathbf{0}, \end{cases}$$

and for all $x_0 \in \mathbb{R}^p$

$$(LP2(x_0)) \qquad \begin{cases} t_2(x_0) = \min \lambda c, \\ \lambda A = x_0^{\mathrm{T}}, \\ \lambda \geqslant \mathbf{0}, \end{cases}$$

with variable $\lambda \in \mathbb{R}^s$, which is a row vector.

We claim that LP can be solved in two phases, where Phase 1 consists in solving $LP1$ and Phase 2 in solving $LP2(x)$ with x resulting from an optimal solution obtained in Phase 1:

Theorem 2.1. *Let $x \in \mathbb{R}^p$ be a column vector, $t \in \mathbb{R}$, and $\lambda \in \mathbb{R}^s$ be a row vector. Statements* (i) *and* (ii) *imply* (iii), *and* (iii) *implies* (i), *where*:
(i) (x, t) *is an optimal solution of* $LP1$,
(ii) (λ) *is an optimal solution of* $LP2(x)$,
(iii) (x, t, λ) *is an optimal solution of* LP.

For the proof of this theorem we shall need the following result of the classical theory of polarity.

Recall that for $X \subset \mathbb{R}^n$, the *polar* X^* of X is the set

$$X^* := \{z \in \mathbb{R}^n \mid z^{\mathrm{T}}x \leqslant 1 \text{ for all } x \in X\}.$$

Proposition 2.2 [11]. *Let $P \subset \mathbb{R}^n$ be a polyhedron with $\mathbf{0} \in P$. Then*:
(i) P^* *is a polyhedron.*
(ii) $P^{**} = P$.
(iii) *If* $P = Conv\{\mathbf{0}, x_1, \ldots, x_m\} + Cone\{y_1, \ldots, y_k\}$ *then* $P^* = \{z \in \mathbb{R}^n \mid x_i^{\mathrm{T}} z \leqslant 1,$ $i = 1, \ldots, m,\ y_j^{\mathrm{T}} z \leqslant 0,\ j = 1, \ldots, k\}$ *and conversely.*

Notice that for the polar P^* of $P = Conv(\{1/c_i \cdot A_i \mid 1 \leqslant i \leqslant s\} \cup \{\mathbf{0}\})$ we then have

$$P^* = Conv(\bar{A}, \mathbf{0}) + Cone(\bar{R}) = \{x \in \mathbb{R}^p \mid 1/c_i \cdot A_i x \leqslant 1,\ 1 \leqslant i \leqslant s\}. \tag{2.2}$$

Proof of Theorem 2.1. We first show that:

(a) If (x', t') is an optimal solution of $LP1$, then $LP2(x')$ has an optimal solution and $t_1 \geqslant t_2(x')$:

If $t_1 = 0$ then $\mu x' \in P$, $\forall \mu \geqslant 0$, hence $x' = \mathbf{0}$ (for P is bounded by (2.1)) and the row vector $\lambda'' := \mathbf{0}$ is an optimal solution of $LP2(x')$ implying $t_2(x') = 0 = t_1$.

If $t_1 \neq 0$ then $y := x'/t_1 \in P$ and by (2.1) there exists $\tilde{\lambda}_i \geqslant 0$, $1 \leqslant i \leqslant s$, such that $y^{\mathrm{T}} = \sum_{i=1}^{s} \tilde{\lambda}_i \cdot 1/c_i \cdot A_i$ and $\sum_{i=1}^{s} \tilde{\lambda}_i \leqslant 1$. Then, the row vector λ'' with $\lambda_i'' := t_1 \cdot \tilde{\lambda}_i / c_i$, $1 \leqslant i \leqslant s$, is feasible for $LP2(x')$. Hence $LP2(x')$ has an optimal solution and $t_2(x') \leqslant \lambda'' c \leqslant t_1$.

We show next that:

(b) If (x, t, λ) is an optimal solution of LP, then (x, t) is feasible for $LP1$:

(i) $\bar{A}_j x = \bar{A}_j(\sum_{i=1}^{s} \lambda_i A_i^{\mathrm{T}}) = \sum_{i=1}^{s} \lambda_i c_i \cdot (\bar{A}_j (1/c_i \cdot A_i)^{\mathrm{T}}) \leqslant \sum_{i=1}^{s} \lambda_i c_i \cdot 1 \leqslant t$ for all rows $\bar{A}_j$ of $\bar{A}$.

(ii) $\bar{R}_j x = \bar{R}_j(\sum_{i=1}^{s} \lambda_i A_i^{\mathrm{T}}) = \sum_{i=1}^{s} \lambda_i c_i \cdot (\bar{R}_j (1/c_i \cdot A_i)^{\mathrm{T}}) \leqslant \sum_{i=1}^{s} \lambda_i c_i \cdot 0 = 0$ for all rows $\bar{R}_j$ of $\bar{R}$.

Now, if (x', t') is an optimal solution of $LP1$ and λ' is an optimal solution of $LP2(x')$, then by (a) (x', t', λ') is a feasible solution of LP. Hence LP has an optimal solution, say (x, t, λ), and $t_0 \leqslant t_1$. By (b), (x, t) is feasible for $LP1$, hence $t_1 \leqslant t_0$. Thus $t_1 = t_0$ and (x', t', λ') is an optimal solution of LP.

Conversely, if (x, t, λ) is an optimal solution of LP, then by (b) (x, t) is feasible for $LP1$. Hence $LP1$ has an optimal solution, say (x', t'), and $t_1 \leqslant t_0$. Moreover, by (a) $LP2(x')$ has an optimal solution λ' and $t_2(x') \leqslant t_1$. Therefore (x', t', λ') is feasible for LP implying $t_0 \leqslant t_1$. Thus $t_0 = t_1$ and (x, t) is an optimal solution of $LP1$. □

As a complement to Theorem 2.1 we conclude this section by giving a geometrical interpretation of the two-phase approach. The main step herein is the following fact:

Proposition 2.3.

$$\mathscr{U} := \{x \in \mathbb{R}^p \mid t_2(x) \leqslant 1\} = \{x \in \mathbb{R}^p \mid \bar{A}x \leqslant \mathbf{1},\ \bar{R}x \leqslant \mathbf{0}\}$$
$$= Conv(\{1/c_i \cdot A_i \mid 1 \leqslant i \leqslant s\} \cup \{\mathbf{0}\}).$$

Proof. Notice that $t_2(x) = \min\{\lambda c \mid \lambda \geqslant \mathbf{0},\ \lambda A = x^{\mathrm{T}}\} = \max\{x^{\mathrm{T}} z \mid Az \leqslant c\}$. It follows that $t_2(x) \leqslant 1 \Leftrightarrow x^{\mathrm{T}} z \leqslant 1\ \forall z \in \{z \in \mathbb{R}^p \mid 1/c_i \cdot A_i z \leqslant 1,\ 1 \leqslant i \leqslant s\}$ so that $\mathscr{U}$ is the polar polyhedron of $\{z \in \mathbb{R}^p \mid 1/c_i \cdot A_i z \leqslant 1,\ 1 \leqslant i \leqslant s\} = P^*$ (see (2.2)), hence by Proposition 2.2(ii) $\mathscr{U}$ is equal to P of (2.1). □

It follows immediately that for $t \in \mathbb{R}_+$, $t\mathscr{U} = \{x \in \mathbb{R}^p \mid \bar{A}x \leqslant t\mathbf{1},\ \bar{R}x \leqslant \mathbf{0}\} = \{x \in \mathbb{R}^p \mid t_2(x) \leqslant t\}$ for $t_2(tx) = t \cdot t_2(x)$. The meaning of this result is that $t\mathscr{U}$ is the set of those points $x \in \mathbb{R}^p$ for which a partition $\lambda A = x^{\mathrm{T}}$, $\lambda \geqslant \mathbf{0}$, can be achieved at a cost $t_2(x) = \lambda c \leqslant t$.

Determining the values t and x of an optimal solution of LP can therefore be achieved by finding the smallest $t \in \mathbb{R}_+$ and $x \in t\mathscr{U}$ such that $Bx + td \leqslant b$, which is actually the object of Phase 1.

Consider now $\beta : \mathbb{R}^p \to \mathbb{R}^n$ defined by $\beta(x) := Bx + d$ and $\mathscr{B} := \{y \in \mathbb{R}^n \mid y \leqslant b\}$. The cost of an optimal solution is the smallest t for which $x = tx' \in t\mathscr{U}$ exists with $Bx + td = t(Bx' + d) = t\beta(x') \leqslant b$ or equivalently for which $t\beta(\mathscr{U}) \cap \mathscr{B} \neq \emptyset$. Hence, Phase 1 consists simply in determining the smallest homothetic ratio $t \in \mathbb{R}_+$ such that $t\beta(\mathscr{U})$ intersects $\mathscr{B}$. If $x' \in \mathscr{U}$ is a point with $t\beta(x') \in \mathscr{B}$ then (t, tx') is an optimal solution of $LP1$. Moreover, any representation λ' of x' as a convex combination of the vertices of $\mathscr{U}$ (i.e. the "weighted operating modes" $1/c_i \cdot A_i$, $1 \leqslant i \leqslant s$, and $\mathbf{0}$) yields an optimal schedule λ with $\lambda_i := t\lambda'_i$, $1 \leqslant i \leqslant s$.

The interpretation of Phase 1 is simpler when $b^{\mathrm{T}}d = 0$, a situation to be considered in the context of scheduling problems with resource constraints at the end of Section 3. In this case $Bx + td \leqslant b$ reduces to the form

$$B'x \leqslant b', \tag{2.3}$$

$$B''x \leqslant td'' \tag{2.4}$$

for some B', b', B'' and d''.

Let $\mathscr{B}' := \{x \in \mathbb{R}^p \mid B'x \leqslant b'\}$ and $\mathscr{B}'' := \{x \in \mathbb{R}^p \mid B''x \leqslant d''\}$. Phase 1 consists then in finding the smallest $t \in \mathbb{R}_+$ and $x \in t\mathscr{U}$ with $B'x \leqslant b'$, $B''x \leqslant td''$, or equivalently with $x \in \mathscr{B}'$ and $x \in t\mathscr{B}''$. Hence Phase 1 can be interpreted in $\mathbb{R}^p$ as determining the smallest homothetic ratio $t \in \mathbb{R}_+$ such that $t(\mathscr{U} \cap \mathscr{B}'')$ intersects $\mathscr{B}'$. As in the general case any representation λ' of x' with $tx' \in t(\mathscr{U} \cap \mathscr{B}'') \cap \mathscr{B}'$ as a convex combination of extremal points of $\mathscr{U}$ yields a schedule $\lambda := t\lambda'$ of cost less than or equal to t (see Fig. 1).

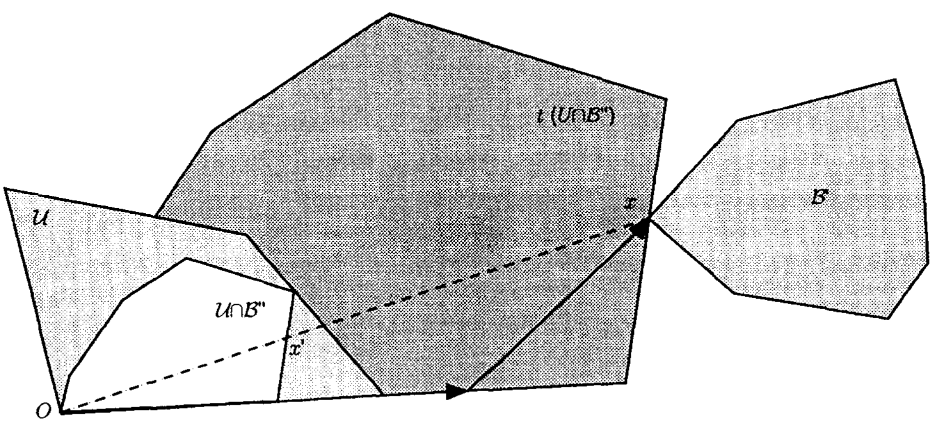

Fig. 1. The cost t of an optimal schedule is equal to the smallest homothetic ratio $t \in \mathbb{R}_+$ such that $t(\mathscr{U} \cap \mathscr{B}'')$ intersects $\mathscr{B}'$.

3. Relations to previous models and resources

Throughout this section we shall consider exclusively minimal length preemptive scheduling problems (i.e. $c = \mathbf{1}$). We first show how the two-phase methods of the previously considered models in [10, 13, 12, 5] fit in our setting, and then discuss the introduction of resource constraints.

Notice first that if the family $\mathscr{F}$ of operating modes is an independence system (i.e. $U \subset V \in \mathscr{F} \Rightarrow U \in \mathscr{F}$), programs LP, $LP1$ and $LP2$ can be given in a simpler form. For A, it suffices to consider the incidence matrix of those operating modes maximal for set inclusion and to replace $\lambda A = x^{\mathrm{T}}$ by $\lambda A \geqslant x^{\mathrm{T}}$ in LP and $LP2$. Furthermore. $\bar{R}x' \leqslant \mathbf{0}$ reduces to $x' \geqslant \mathbf{0}$.

The situations considered in [10, 13, 12, 5] fit in this simplified model. In the basic model of [10, 13] Phase 1 consists in solving $LP1''$ of Section 1, which is equal to $LP1$ since $\bar{A}$ is given by inequalities (1.1) and (1.2) (Birkhoff–Von Neumann theorem in bistochastic matrices). The families of operating modes considered in [12] and generalized in [5] are subsets of the operating modes of the basic model. More precisely, a nonintersecting family $\mathscr{P}$ of subsets of the processors $\boldsymbol{P}$ is given (i.e. U, $V \in \mathscr{P}$ and $U \cap V \neq \emptyset \Rightarrow U \subset V$ or $V \subset U$) with a bound $\gamma(U)$ associated to each subset $U \in \mathscr{P}$. The family $\mathscr{F}$ of operating modes in [12] is obtained from the family of the basic model by excluding those operating modes which require more than $\gamma(U)$ processors for some subset $U \in \mathscr{P}$. In fact this condition expresses a limitation on the number of processors which can work simultaneously. The family $\mathscr{G}$ of operating modes in [5] is obtained from $\mathscr{F}$ by additionally imposing similar conditions on the jobs, by means of another nonintersecting family. For the models of [12, 5] the authors showed the existence of a two-phase method, in particular they exhibited a linear program for Phase 1. This linear program is equivalent to our $LP1$, as shown in [1] by the following argument. The family of those subsets of $\boldsymbol{P} \times \boldsymbol{J}$ satisfying the conditions given by $\mathscr{P}$ and γ is actually the family of all independent sets of a matroid, so that the family $\mathscr{G}$ is the family of all sets independent in two matroids. The polyhedral description of $\mathscr{G}$ then follows from a well-known result of Edmonds and, once inserted in $LP1$, delivers a linear program equivalent to the one of [5].

To conclude this section we discuss how the preemptive scheduling problem LP' considered in Section 1 can be extended to include some additional resource constraints, while remaining in the setting of LP. Recall that in LP' $Bx + td \leqslant b$ reduces to (2.3) so that the addition of resource constraints will be illustrated geometrically on the basis of Fig. 1.

Let us first consider renewable resources. In this case a bound is given for each renewable resource K and the amount of K available at any instant is limited by this bound. Notice that the need for such a resource at a given instant is completely determined by the operating mode in application at this instant. Taking into account renewable resources consequently amounts to excluding those operating modes which consume too much of one of them and geometrically corresponds to a modification of the polyhedron $\mathscr{U}$ (Proposition 2.3). This situation clearly remains in the setting of our

LP. As an example, the model given in [5] and recalled just above has originally been presented as the basic model of [10, 13] together with renewable resource constraints: the processors (or jobs) belonging to some subsets U of the given nonintersecting family are considerd as a renewable resource, the availability of which is limited at any time by $\gamma(U)$.

We now consider models involving nonrenewable resources. For any resource K of this type the amount k_{ij} of K required to process job J_j on processor P_i for one time unit is given and the amount of K available up to time τ is limited for any τ.

Let us discuss two cases according to the availability of resources:

(a) If for a given resource K the total amount k at disposal is bounded and already available at the beginning of the schedule this resource can be taken into account in the model LP: the constraint $\sum x_{ij}k_{ij} \leqslant k$ has to be added to the system $Bx \leqslant b$ of the preemptive scheduling problem LP'. Geometrically this constraint corresponds to a cut of the polyhedron $\mathscr{B}'$ (see Fig. 1). This cut excludes those points in $\mathscr{B}'$ which cannot be reached by a lack of resource K.

(b) For resources K becoming available at a linear rate of ρ_K units per time unit, the order of application of the operating modes does matter and, conceptually, the situation is beyond the scope of linear programming. However, following an argument of [2], the problem with a unique nonrenewable resource K with rate ρ can be handled within model LP, provided that $\mathbf{0}$ is an operating mode (i.e. doing nothing is allowed).

The necessary condition

$$\sum x_{ij}k_{ij} \leqslant \rho t \tag{3.1}$$

has to be introduced in LP', yielding again a problem of type LP. This LP solves the scheduling problem in the following sense. Assume that the operating modes are numbered in order of increasing consumption of K, that is $\sum_j A_{ij}k_{ij} \leqslant \sum_j A_{(i+1)j}k_{(i+1)j}$, $1 \leqslant i < s$. Let (x, t, λ) be an optimal solution of LP with t equal to $\lambda\mathbf{1} \equiv \lambda c$ (since $\mathbf{0}$ is an operating mode, such a solution exists if LP has a solution). An optimal schedule respecting the availability of K is then obtained by applying the operating modes in the order $1, \ldots, s$. The introduction in LP' of the constraint (3.1) geometrically corresponds to a cut of the polyhedron $\mathscr{U} \cap \mathscr{B}''$. This cut excludes those points which cannot be reached in one time unit, because the availability of resource K is too slow.

4. Almost nonpreemptive schedules

In this section we shall briefly examine the topic of almost nonpreemptive scheduling as introduced in [7]. As a result of Phase 1, we may for each job J_j have several processors P_i for which $x_{ij} > 0$. This introduces preemptions in the processing of job J_j called first-order preemptions. The partition obtained in Phase 2 will in general give schedules where J_j will not be on P_i for x_{ij} consecutive time units. These interruptions are called second-order preemptions. As a consequence of such preemptions, a job may be processed more than once on a given processor. A schedule with no second-order preemption is said to be almost nonpreemptive.

In order to get an almost nonpreemptive schedule by the two-phase method, the submatrix of A corresponding to the support of the solution λ of Phase 2 must have the so-called consecutive 1 property [8].

We present now a class of minimal length almost nonpreemptive scheduling problems (i.e. $c = \mathbf{1}$) which can be optimally solved by the two-phase method. This class is characterized by the matrix of operating modes being a so-called lattice matrix [4]. Recall that a $\{0, 1\}$-matrix A is a lattice matrix if its rows can be indexed by the elements of a distributive lattice $\mathscr{L}$ and the columns f_j of A define nonzero $\{0, 1\}$-valued functions on $\mathscr{L}$ which are consecutive ($a < b < c$ and $f_j(a) = f_j(c) = 1$ imply $f_j(b) = 1$), modular ($f_j(a \wedge b) + f_j(a \vee b) = f_j(a) + f_j(b)$), and such that $f_j(m) = f_j(M) = 0$ for m (resp. M) being the minimal (resp. maximal) element of $\mathscr{L}$.

An example of a lattice matrix is given by the incidence matrix of all convex sets in a poset $\mathscr{H}$ ($U \subset \mathscr{H}$ is a convex set if $a, c \in U$ and $a < b < c$ imply $b \in U$). If $\mathscr{H}$ is the set of actions $\boldsymbol{P} \times \boldsymbol{J}$ (i.e. a partial order is given on the actions), the convexity condition says that when actions a and b with $a < b$ occur simultaneously (i.e. belong to the same operating mode), then all actions on chains between a and b must occur too: they are "byproducts" of a and b. This model can be improved by considering for instance only those convex sets whose maximal chain has cardinality bounded by a given constant. The corresponding matrix of operating modes is still a lattice matrix [4].

If A is a lattice matrix, it is known from [9] that whenever $LP2(x)$ has a solution, it has an optimal solution whose support corresponds to a submatrix of A having the consecutive 1 property. Hence there exists a two-phase method for the minimal length almost nonpreemptive scheduling problem. Moreover, all elements for applying this method are available from [3, 4]. The polyhedral description of (the rows of) lattice matrices is known as well as a partitioning algorithm for Phase 2, which is based on a longest path computation in a digraph. Moreover, this algorithm delivers an ordering of the support of its solution, yielding an almost nonpreemptive schedule.

To sum up, if A is a lattice matrix, then an almost nonpreemptive schedule always exists with length equal to that of a minimal length preemptive schedule, and it can be determined by a two-phase approach.

Acknowledgements

The authors are grateful to an anonymous referee for pointing out the possibility of introducing a nonuniform cost vector and for valuable comments in relation with the exposition of the paper.

References

[1] M. Cochand, Sur quelques problèmes combinatoires relatifs à l'ordonnancement, Thèse E.P.F.L No. 789 (1990).

[2] M. Cochand, D. de Werra and R. Slowinski, Preemptive scheduling with staircase and piecewise linear resource availability, ZOR-Methods Models Oper. Res. 33 (1989) 297–313.

[3] M. Cochand, A. Gaillard and H. Gröflin, Optimum intersection of two ring families, Res. Report IFOR-Studienberichte 23, Zurich (1988).

[4] M. Cochand, A. Gaillard and H. Gröflin, Lattice matrices, intersection of ring families and dicuts, Discrete Math. 110 (1992) 61–80.

[5] D. de Werra, Preemptive scheduling, linear programming and network flows, SIAM J. Algebraic Discrete Methods 5 (1984) 11–20.

[6] D. de Werra, On the two-phase method for preemptive scheduling, European J. Oper. Res. 37 (1988) 227–235.

[7] D. de Werra, Almost nonpreemptive schedules, Ann. Oper. Res. 26 (1990) 243–256.

[8] D.R. Fulkerson and O.A. Gross, Incidence matrices and interval graphs, Pacific J. Math. 15 (1965) 835–855.

[9] A.J. Hoffman and D.E. Schwartz, On lattice polyhedra, in: A. Hajnal and V.T. Sós, eds., Combinatorics (North-Holland, Amsterdam, 1978) 593–398.

[10] E.L. Lawler and J. Labetoulle, On preemptive scheduling of unrelated processors by linear programming, J. ACM 25 (1978) 612–619.

[11] A. Schrijver, Theory of Linear and Integer Programming (Wiley, Chichester, 1986).

[12] R. Slowinski, L'ordonnancement des tâches préemptives sur les processeurs indépendants en présence de resources supplémentaires, RAIRO Inform. 15 (1981) 155–166.

[13] R. Slowinski and J. Weglarz, Modèle d'ordonnancement optimal avec plusieurs modes d'exécution des opérations, Przeglad Statyst. 24 (1977) 409–416.